Numer. Math. 20, 93—98 (1972)

On Asymptotically Best Norms for Powers of Operators

W. J. Kammerer and R. S. Varga*

Received May 5, 1972

Abstract. In the study of the successive overrelaxation iterative method for solving large systems of linear equations, a frequently considered problem is the behavior of the norm of powers of the successive overrelaxation matrix, $\mathcal{L}_{\omega_b}^m$, both as a function of m and of the given norm. Our main result is a rather natural necessary and sufficient condition for the existence of a norm asymptotically best for a non-nilpotent matrix A. As a corollary of our main result, it is shown here that there is *no* asymptotically best norm for the successive overrelaxation matrix $\mathcal{L}_{\omega_b}$.

1. Introduction

If $[\mathbb{C}^n]$ denotes the collection of all $n \times n$ complex matrices, and if $\mathcal{F}$ denotes the collection of all vector norms on $\mathbb{C}^n$, the vector space of all column vectors with n components, then as usual,

$$\|A\|_\phi \equiv \sup\{\phi(Ax) : \phi(x) = 1\}, \qquad A \in [\mathbb{C}^n], \qquad \phi \in \mathcal{F},$$

denotes the induced operator norm of A with respect to ϕ, and

$$\varrho(A) \equiv \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$$

denotes the spectral radius of A. Then, given $A \in [\mathbb{C}^n]$ and given $\phi \in \mathcal{F}$, the sequence $\{\|A^m\|_\phi\}_{m=1}^\infty$ can be formed, and of specific interest here is the asymptotic behavior, as $m \to \infty$, of $\|A^m\|_\phi$.

Definition 1. A norm $\phi \in \mathcal{F}$ is *asymptotically best* for $A \in [\mathbb{C}^n]$ if, for every $\psi \in \mathcal{F}$, there exists a finite positive integer $m(\psi)$ for which

$$\|A^m\|_\phi \leqq \|A^m\|_\psi \quad \text{for all} \quad m \geqq m(\psi). \tag{1.1}$$

We remark that if $\varrho(A) = 0$, i.e., $A \in [\mathbb{C}^n]$ is nilpotent, then it is easily seen, as a rather trivial case, that *every* norm in $\mathcal{F}$ is asymptotically best for A.

Before proving our main result, we need to recall the following definition, introduced by Householder [1].

Definition 2. A matrix $A \in [\mathbb{C}^n]$ is of *class M* if there exists a $\phi \in \mathcal{F}$ for which

$$\|A\|_\phi = \varrho(A). \tag{1.2}$$

Equivalently (cf. Householder [2, p. 47]), $A \in [\mathbb{C}^n]$ is of class M if and only if, for every eigenvalue λ of A with $|\lambda| = \varrho(A)$, the number of linearly independent

* Research supported in part by the Atomic Energy Commission under Grant AT(11-1)-2075.

eigenvectors associated with λ is equal to the multiplicity of λ. In other words, the Jordan blocks, in the Jordan normal form for A, corresponding to those eigenvalues λ of A with $|\lambda| = \varrho(A)$ are all 1×1 matrices.

2. Main Result

With the above definitions, we now prove

Theorem 1. For any $A \in [\mathbb{C}^n]$ with $\varrho(A) > 0$, there exists an asymptotically best norm $\phi \in \mathscr{F}$ for A if and only if A is of class M.

Proof. First, assume that A is of class M, so that there exists a $\phi \in \mathscr{F}$ for which (1.2) is valid. Then, for any positive integer m,

$$\varrho^m(A) = \varrho(A^m) \leqq \|A^m\|_\phi \leqq \|A\|_\phi^m = \varrho^m(A),$$

i.e.,

$$\varrho(A^m) = \|A^m\|_\phi \quad \text{for all} \quad m = 1, 2, \ldots.$$

Thus, for any $\psi \in \mathscr{F}$, it follows that

$$\|A^m\|_\phi = \varrho(A^m) \leqq \|A^m\|_\psi \quad \text{for all} \quad m = 1, 2, \ldots,$$

showing that ϕ is asymptotically best for A.

Conversely, assuming that A is neither nilpotent nor of class M, we shall show that there is *no* asymptotically best norm for A. More precisely, we shall show that, given *any* $\psi \in \mathscr{F}$, we can construct a norm $\phi \in \mathscr{F}$ for which (cf. (1.1))

$$\|A^m\|_\phi < \|A^m\|_\psi \quad \text{for all } m \text{ sufficiently large.} \tag{2.1}$$

To do this, consider first the special case where $A \in [\mathbb{C}^n]$, $n > 1$, with

$$A = \lambda I + N, \qquad \lambda \neq 0, \tag{2.2}$$

where N is the strictly lower triangular matrix given by $N = [\boldsymbol{e}_2, \boldsymbol{e}_3, \ldots, \boldsymbol{e}_n, \boldsymbol{0}]$, with $\boldsymbol{e}_i$ denoting i-th column vector of the $n \times n$ identity matrix. Then, given any $\psi \in \mathscr{F}$, let $K_\psi = \{\boldsymbol{x} \in \mathbb{C}^n \colon \psi(\boldsymbol{x}) \leqq 1\}$ denote its unit ball. Since K_ψ and αK_ψ, $\alpha > 0$, induce the same operator norm, we may assume that K_ψ is normalized so that $K_\psi \subset K_\infty \equiv \left\{\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i \in \mathbb{C}^n \colon \|\boldsymbol{x}\|_\infty \equiv \max_{1 \leqq i \leqq n} |x_i| \leqq 1\right\}$. Let $\boldsymbol{a} = a\boldsymbol{e}_1$, $a > 0$, be the vector for which $\psi(\boldsymbol{a}) = 1$. Since $K_\psi \subset K_\infty$, then evidently $0 < a \leqq 1$. Now, choose $\varepsilon > 0$ such that $0 < \varepsilon^{n-1} < a \leqq 1$. Next, we recall (cf. Householder [2, p. 41]) that for any nonsingular $P \in [\mathbb{C}^n]$ and for any $\sigma \in \mathscr{F}$ with unit ball K_σ, then $K_\tau = PK_\sigma \equiv \{\boldsymbol{y} \in \mathbb{C}^n \colon \boldsymbol{y} = P\boldsymbol{x} \text{ for some } \boldsymbol{x} \in K_\sigma\}$ is the unit ball for a norm $\tau \in \mathscr{F}$, and moreover, as $\tau(\boldsymbol{x}) = \sigma(P^{-1}\boldsymbol{x})$, then

$$\|A\|_\tau = \|P^{-1} A P\|_\sigma. \tag{2.3}$$

In particular, defining $W \in [\mathbb{C}^n]$ by $W \equiv \operatorname{diag}\{1, \varepsilon, \varepsilon^2, \ldots, \varepsilon^{n-1}\}$, then as W is nonsingular, $K_\phi \equiv W^{-1} K_\infty$ defines the unit ball for a norm $\phi \in \mathscr{F}$. We see from these definitions that $K_\psi \subset K_\infty \subset K_\phi$, so that

$$\phi(\boldsymbol{x}) \leqq \|\boldsymbol{x}\|_\infty \leqq \psi(\boldsymbol{x}) \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{C}^n. \tag{2.4}$$

Next, from (2.2), we see that

$$A^m = (\lambda I + N)^m = \sum_{k=0}^{n-1} \binom{m}{k} \lambda^{m-k} N^k, \qquad m \geqq n-1,$$

and applying (2.3) simply gives

$$\|A^m\|_\phi = \|W A^m W^{-1}\|_\infty = \sum_{k=0}^{n-1} \binom{m}{k} \varepsilon^k |\lambda|^{m-k}. \tag{2.5}$$

On the other hand, since $\|A^m\|_\psi = \sup\{\psi(A^m \boldsymbol{x}) \colon \psi(\boldsymbol{x}) = 1\}$, and since $\psi(\boldsymbol{a}) = 1$, then

$$\|A^m\|_\psi \geqq \psi(A^m \boldsymbol{a}) \geqq \|A^m \boldsymbol{a}\|_\infty, \tag{2.6}$$

the last inequality following from (2.4). A short calculation shows from the definition of $\boldsymbol{a}$ that $\|A^m \boldsymbol{a}\|_\infty = a \binom{m}{n-1} |\lambda|^{m-n+1}$ for all m sufficiently large, i.e., from (2.6),

$$\|A^m\|_\psi \geqq a \binom{m}{n-1} |\lambda|^{m-n+1} \quad \text{for all } m \text{ sufficiently large.} \tag{2.7}$$

Thus, as $0 < \varepsilon^{n-1} < a$, a comparison of (2.5) and (2.7) gives us that (cf. (2.1))

$$\|A^m\|_\phi < \|A^m\|_\psi \quad \text{for all } m \text{ sufficiently large.} \tag{2.8}$$

For the general case, assume now that $A \in [\mathbb{C}^n]$ is an arbitrary matrix which is neither nilpotent (i.e., $\varrho(A) > 0$), nor of class M. Because of (2.3), there is no loss of generality in assuming that A is in Jordan normal form, i.e., $A = \operatorname{diag}[J_1, J_2, \ldots, J_r]$, where each of the square submatrices J_i is either a 1×1 matrix or a $p_i \times p_i$ matrix, $p_i > 1$, of the form

$$J_i = \begin{bmatrix} \lambda_i & 0 & \cdots & 0 & 0 \\ 1 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & \lambda_i \end{bmatrix}. \tag{2.9}$$

Assuming $A = \operatorname{diag}[J_1, \ldots, J_r]$ means that the vector space $\mathbb{C}^n$ can be decomposed into r pairwise disjoint subspaces W_i, $1 \leqq i \leqq r$, i.e.,

$$\mathbb{C}^n = W_1 \oplus W_2 \oplus \cdots \oplus W_r, \tag{2.10}$$

and each W_i is invariant under A.

Given the decomposition of $\mathbb{C}^n$ in (2.10), let P_i denote the projection mapping of $\mathbb{C}^n$ onto W_i, and let T_i denote the natural injection mapping of W_i into $\mathbb{C}^n$, whereby appropriate zero components are simply added to an $\boldsymbol{x}_i \in W_i$ to define an analogous vector in $\mathbb{C}^n$. Then,

$$\psi_i(\boldsymbol{x}_i) \equiv \psi(T_i \boldsymbol{x}_i), \qquad \boldsymbol{x}_i \in W_i, \qquad 1 \leqq i \leqq r,$$

defines a norm on W_i. Since W_i is a subspace of $\mathbb{C}^n$ and since A on W_i is given by J_i, then for each i,

$$\|A^m\|_\psi \equiv \sup\{\psi(A^m \boldsymbol{x}) \colon \psi(\boldsymbol{x}) = 1\} \geqq \sup\{\psi(A^m T_i \boldsymbol{x}_i) \colon \boldsymbol{x}_i \in W_i \text{ and } \psi(T_i \boldsymbol{x}_i) = 1\}$$
$$= \sup\{\psi_i(J_i^m \boldsymbol{x}_i) \colon \boldsymbol{x}_i \in W_i \text{ and } \psi_i(\boldsymbol{x}_i) = 1\} = \|J_i^m\|_{\psi_i},$$

where $\|\cdot\|_{\psi_i}$ denotes the induced operator norm on W_i with respect to the norm ψ_i. As this inequality is valid for each i, $1 \leqq i \leqq r$, then

$$\|A^m\|_\psi \geqq \max_{1\leqq i\leqq r} \{\|J_i^m\|_{\psi_i}\}. \tag{2.11}$$

We now make use of our previous analysis for matrices of the form (2.2). If J_i is either a 1×1 matrix or J_i is given by (2.9) with $\lambda_i=0$, then define the norm ϕ_i on W_i to be just the L_∞-norm on W_i. If J_i is given by (2.9) with $\lambda_i \neq 0$ and $p_i>1$, we can select the norm ϕ_i on W_i so that (cf. (2.8))

$$\|J_i^m\|_{\phi_i} < \|J_i^m\|_{\psi_i} \quad \text{for all } m \text{ sufficiently large.} \tag{2.12}$$

With these norms ϕ_i on W_i, $1\leqq i\leqq r$, we then define the norm ϕ on $\mathbb{C}^n$ by

$$\phi(\boldsymbol{x}) \equiv \max_{1\leqq i\leqq r} \phi_i(P_i \boldsymbol{x}) \quad \text{for all} \quad \boldsymbol{x}\in\mathbb{C}^n. \tag{2.13}$$

Then, since A is assumed to be neither nilpotent or of class M, there is at least one J_i from $A=\operatorname{diag}[J_1, J_2, \ldots, J_r]$ for which (cf. (2.9)) $0<|\lambda_i|=\varrho(A)$ and for which $p_i>1$. It then easily follows from (2.11) and (2.12) that the norm ϕ so defined on $\mathbb{C}^n$ by (2.13) is such that

$$\|A^m\|_\psi > \|A^m\|_\phi \quad \text{for all } m \text{ sufficiently large,}$$

which shows that no norm on $\mathbb{C}^n$ is asymptotically best for A. Q.E.D.

The following corollaries are direct consequences of Theorem 1 and its proof.

Corollary 1. If $A\in[\mathbb{C}^n]$ is of class M, then there exists a norm $\phi\in\mathscr{F}$ for which

$$\varrho(A^m) = \|A^m\|_\phi = \|A\|_\phi^m \quad \text{for all} \quad m=1, 2, \ldots. \tag{2.14}$$

Corollary 2. If p is any complex polynomial and if $A\in[\mathbb{C}^n]$ is similar to a diagonal matrix, then there exists a norm $\phi\in\mathscr{F}$ which is asymptotically best for $p(A)$.

Corollary 3. If $A\in[\mathbb{C}^n]$ is neither nilpotent nor of class M, then for any $\psi\in\mathscr{F}$, there exists a $\phi\in\mathscr{F}$ for which

$$\|A^m\|_\phi < \|A^m\|_\psi \quad \text{for all } m \text{ sufficiently large.} \tag{2.15}$$

If $A\in[\mathbb{C}^n]$, $n>1$, is of class M, we remark that there are infinitely many norms $\phi\in\mathscr{F}$ for which (2.14) is valid.

3. Application

If $G\in[\mathbb{C}^n]$ is a positive definite Hermitian matrix of the form

$$G = \begin{bmatrix} I_1 & -B_1^* \\ -B_1 & I_2 \end{bmatrix} \equiv I - B, \tag{3.1}$$

where I_1 and I_2 are respectively $r\times r$ and $(n-r)\times(n-r)$ identity matrices, $1<r<n$, then the well-known successive overrelaxation matrix $\mathscr{L}_\omega$ is defined by

$$\mathscr{L}_\omega = (I-\omega L)^{-1}\{\omega L^* + (1-\omega) I\}, \tag{3.2}$$

where ω is the relaxation factor and where $L \in [\mathbb{C}^n]$ is the strictly lower triangular matrix determined from (3.1) by

$$L = \begin{bmatrix} 0 & 0 \\ B_1 & 0 \end{bmatrix}.$$

It is also well known that

$$\min_{\omega \text{ real}} \varrho(\mathscr{L}_\omega) = \varrho(\mathscr{L}_{\omega_b}) = \omega_b - 1, \tag{3.3}$$

where $\omega_b = \dfrac{2}{1+\sqrt{1-\varrho^2(B)}}$, with $B \in [\mathbb{C}^n]$ being defined from G in (3.1).

Of late, there has been renewed interest in the behavior of the norms of $\mathscr{L}_{\omega_b}^m$ as $m \to \infty$. In particular, it is known (cf. Young [7, p. 248]) that

$$\|\mathscr{L}_{\omega_b}^m\|_{\phi_1} = r^m \{m(r^{-\frac{1}{2}} + r^{\frac{1}{2}}) + [m^2(r^{-\frac{1}{2}} + r^{\frac{1}{2}})^2 + 1]^{\frac{1}{2}}\}, \quad m = 1, 2, \ldots, \tag{3.4}$$

where $r \equiv \omega_b - 1 = \varrho(\mathscr{L}_{\omega_b})$, and where $\phi_1(\boldsymbol{x}) \equiv (x^* x)^{\frac{1}{2}}$ is the usual L_2-norm on $\mathbb{C}^n$. Similarly, it has been recently shown (cf. Young [7, p. 258] and Young and Kincaid [8]) that

$$\|\mathscr{L}_{\omega_b}^m\|_{\phi_2} = r^m \{m(r^{-\frac{1}{2}} - r^{\frac{1}{2}}) + [m^2(r^{-\frac{1}{2}} - r^{\frac{1}{2}})^2 + 1]^{\frac{1}{2}}\}, \quad m = 1, 2, \ldots, \tag{3.5}$$

where again $r = \omega_b - 1 = \varrho(\mathscr{L}_{\omega_b})$, and where $\phi_2(\boldsymbol{x}) \equiv \phi_1(G^{\frac{1}{2}}\boldsymbol{x}) = (x^* G x)^{\frac{1}{2}}$ is a norm on $\mathbb{C}^n$, since G is by hypothesis a positive definite Hermitian matrix. It is clear from (3.4) and (3.5) that

$$\|\mathscr{L}_{\omega_b}^m\|_{\phi_2} < \|\mathscr{L}_{\omega_b}^m\|_{\phi_1} \quad \text{for all} \quad m = 1, 2, \ldots.$$

These above investigations, as well as those of others (cf. Kincaid [3, 4, and 5]), could be interpreted as having, in addition to other objectives, the goal of finding norms which are asymptotically best for $\mathscr{L}_{\omega_b}$. As the next corollary shows, this goal simply cannot be attained.

Corollary 4. Given the matrix G of (3.1), assume that G is Hermitian and positive definite, and that $\varrho(B) > 0$. Then, for any $\psi \in \mathscr{F}$, there is a $\phi \in \mathscr{F}$ for which

$$\|\mathscr{L}_{\omega_b}^m\|_{\phi} < \|\mathscr{L}_{\omega_b}^m\|_{\psi} \quad \text{for all } m \text{ sufficiently large.} \tag{3.6}$$

Proof. The assumption that $\varrho(B) > 0$ implies from (3.3) that $\mathscr{L}_{\omega_b}$ is not nilpotent. Further, as it is known (cf. Varga [6, p. 111] and Young [7, p. 238]) that $\mathscr{L}_{\omega_b}$ is not of class M, then the inequality of (3.6) follows directly from Corollary 3. Q.E.D.

We finally remark that the result of Corollary 4 is also valid for the more general case of the block successive overrelaxation method in which the matrix A is a consistently ordered p-cyclic matrix, $p \geqq 2$, for which the eigenvalues of B^p, B being the associated Jacobi matrix, are assumed to be real, and nonnegative, with $0 < \varrho(B) < 1$. The proof analogously depends on the fact that the matrix $\mathscr{L}_{\omega_b}$ is again neither nilpotent or of class M (cf. Varga [6, p. 111]).

Bibliography

1. Householder, A. S.: Minimal matrix norms. Monatsh. Math. **63**, 344–350 (1959).
2. Householder, A. S.: The theory of matrices in numerical analysis. New York: Blaisdell Publishing Co. 1964 (257 pp.).
3. Kincaid, D. R.: An analysis of a class of norms of iterative methods for system of linear equations. Doctoral thesis, The University of Texas at Austin, May 1971 (193 pp.).
4. Kincaid, D. R.: Norms of the successive overrelaxation method. Math. of Comp. **26**, 345–357 (1972).
5. Kincaid, D. R.: A class of norms of iterative methods for solving systems of linear equations. Numer. Math. (to appear).
6. Varga, R. S.: Matrix iterative analysis. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. 1962 (322 pp.).
7. Young, D. M.: Iterative solution of large linear systems. New York: Academic Press, Inc. 1971 (570 pp.).
8. Young, D. M., Kincaid, D. R.: Norms of the successive overrelaxation method and related methods. TNN-94, Computation Center, University of Texas, September, 1969 (101 pp.).

W. J. Kammerer
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332 U.S.A.

Richard S. Varga
Department of Mathematics
Kent State University
Kent, OH 44242 U.S.A.